Audibility of all-pass components in head-related transfer functions

Plogsties, Jan; Olesen, Søren Krarup; Minnaar, Pauli; Christensen, Flemming; Møller, Henrik

Published in:
Proceedings of 108th Audio Engineering Society Convention

Publication date:
2000

Audibility of All-Pass Components in Head-Related Transfer Functions

5132 (L - 1)

Jan Plogsties, Søren Krarup Olesen, Pauli Minnaar, Flemming Christensen and Henrik Møller
Acoustics Laboratory, Aalborg University, Denmark

Presented at the 108th Convention 2000 February 19-22 Paris, France

AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

AN AUDIO ENGINEERING SOCIETY PREPRINT

Audibility of all-pass components in head-related transfer functions

Jan Plogsties, Pauli Minnaar, Søren Krarup Olesen,
Flemming Christensen, Henrik Møller

Acoustics Laboratory, Aalborg University
Fredrik Bajers Vej 7 B4, DK-9220 Aalborg Ø, Denmark
acoustics@kom.auc.dk

Abstract

A measured head-related transfer function can be decomposed into a minimum phase, a linear phase and an all-pass component. The audibility of the all-pass component is tested in a three-alternative forced choice experiment. Results show that a pair of head-related transfer functions can be represented by its minimum phase components if the appropriate interaural time difference is applied. The interaural time difference is found as the interaural group delay difference at 0 Hz determined from the excess phase (linear phase and all-pass) components.

1 Introduction

Binaural synthesis is the process through which binaural signals are created by means of head-related transfer functions (HRTFs). HRTFs are transfer functions that describe the transmission of sound from a point is space to the ears of a person. An HRTF is defined as the sound pressure at the blocked entrance of the ear canal divided (in frequency) by the sound pressure measured at the position of the middle of the head with the listener absent [1]. If an anechoic signal is convolved with a pair of HRTFs and presented to a listener wearing equalized headphones, the sound appears to originate from the particular direction.

All acoustical information about the direction of the sound is contained in measured HRTFs. However, more efficient methods of representing the HRTFs may be found that give the same perception. In the current study HRTFs are decomposed into three components: 1) a minimum phase component, 2) a linear phase component (pure delay) and 3) an all-pass component:

$$H(z) = H(z)_{minimum\ phase} \cdot H(z)_{excess\ phase} \quad (1)$$

$$H(z) = H(z)_{minimum\ phase} \cdot H(z)_{linear\ phase} \cdot H(z)_{all-pass} \quad (2)$$

The aim is to test the audibility of the all-pass components in a subjective experiment. In a three-alternative forced choice test (3-AFC) measured HRTFs are compared to HRTF representations without the all-pass component. It is also examined whether the all-pass component can be exchanged by a pure delay. If the all-pass components could be discarded, binaural synthesis would be much simplified since a sound source can be implemented by a minimum phase function for each ear and an interaural time difference (ITD).

A minimum phase system is the causal, stable system with the smallest possible phase lag, group delay and energy delay for a given magnitude [2]. This implies that minimum phase filters used for representing the HRTFs are the shortest possible. When a moving sound source or listener is required, HRTFs must change systematically with the direction of the sound source. Therefore, interpolations must be made between HRTFs measured at discrete directions. Here, minimum phase representations of neighboring HRTFs are more similar than the original ones.

2 Previous work

The idea of representing the HRTFs by a minimum phase function for each ear and a pure delay as ITD is not new. Since Mehrgardt and Mellert [3] in 1977 claimed that the phase of the all-pass component is nearly linear up to 10 kHz, many authors have implemented HRTFs as minimum phase functions.

Jot et al. [4] described a method for determining the ITD to apply to the minimum phase functions. This delay is estimated by fitting a linear curve to the phase of the excess-phase component between 1 kHz and 5 kHz. The same method is mentioned by Huopaniemi and Smith [5] but here a frequency range of 500 Hz to 2 kHz is suggested. Kistler and Wightman [6] presented a model of HRTFs based on principle component analysis, which assumes that the HRTFs are minimum phase. The ITD is calculated as the delay corresponding to the maximum in the cross-correlation function between a measured HRTF pair. Yet another method for finding the ITD was described by Sandvad and Hammershøi [7]. The sample where the impulse response for the first time exceeds 5% of its maximum value are found for the left and right HRTF and subtracted to obtain an estimate of the ITD. These three methods can give very different ITDs. Detailed descriptions of the different methods for calculating the ITD are given in [8].

Subjective experiments found in the literature generally conclude that minimum phase HRTFs cannot be distinguished from the measured HRTFs except in a few cases. Sandvad and Hammershøi [7] compared 17 measured and minimum phase HRTFs in a three-alternative forced choice experiment with 13 listeners and found that some listeners may be able to detect the difference for some minimum phase functions. Kulkarni et al. [9] compared measured HRTFs to their minimum phase counterparts in a four-interval, two-alternative forced choice listening experiment. Listeners generally had chance performance except for two of the four listeners who had performances that were significantly better than chance for the +90 and -90 degrees (directly left and right) positions in the horizontal plane.

Although minimum phase HRTFs with a pure delay as ITD are now widely applied, some may still be audibly different from measured HRTFs. Criteria for the audibility are not found in the literature and an exhaustive analysis of the differences has not been described.

3 Hypothesis

Minnaar et al. [10] tested the audibility of single second order all-pass sections for diotic and dichotic presentation through headphones. Thresholds for high Q-factors and low Q-factors were determined for both causal and non-causal all-pass sections. It was shown that the 'low Q', causal, dichotic threshold relates to the group delay $\tau_g(f)$ of the all-pass section in a simple way. It was found that the introduction of a second order all-pass section to one channel of the headphone will be audible if $\tau_g(0) > 30\ \mu s$.

Since the all-pass component in an HRTF consists of first and second order all-pass sections (described in more detail in the next section) the group delays of all sections add up to the group delay of the complete all-pass component. Therefore, if an arbitrary number of all-pass sections is applied to each ear, the audibility will be controlled by the interaural group delay difference evaluated at 0 Hz, IGD_0, where

$$IGD_0 = abs\ (\ \tau_g(0)_{left} - \tau_g(0)_{right}\ ).$$

So, the influence of the all-pass components is expected to be audible if the IGD_0 introduced by the all-pass components of an HRTF pair is above the 30 μs threshold. Removing the all-pass components in such a case will be audible. Analysis of HRTFs of 40 people measured at 97 directions [1] revealed that this is possible for a few directions. The minimum phase and all-pass components of an HRTF pair where the absence of the all-pass components will be audible are shown in Figure 1. The corresponding group delay functions of the all-pass components are given in Figure 2.

From Minnaar et al. [10] it could be derived that the all-pass components introduce a perceived interaural delay which is IGD_0. This suggests a method for finding the ITD to use in combination with the minimum phase functions that guarantees no detectability from the measured HRTFs. The delay from each HRTF is determined as the linear phase component plus $\tau_g(0)$ from the all-pass component. The delays of the two sides are subtracted to give the ITD. This is equivalent to specifying the ITD as the IGD_0 of the excess phase components. So the hypothesis is, that minimum phase HRTFs are indistinguishable from measured HRTFs if the ITD is the IGD_0 determined from the excess phase components.

4 Experimental methods

Extensive pilot listening experiments have shown the hypothesis to be true for the HRTFs of 40 people. In order to show this in a formal experiment 'worst case' HRTFs - that give the largest (above threshold) IGD_0 of the all-pass components - were found. It was expected in the subjective experiment that simply removing the all-pass components of these HRTFs would be audible. But if the IGD_0 of the all-pass components is added to the linear phase components, no detection is expected. In addition, typical HRTFs, having a small IGD_0 of the all-pass components, were selected from the database. No detection was expected for these HRTFs when the all-pass components are removed since the IGD_0s of the all-pass components are in the range of 0-30 μs.

4.1 Decomposition of the HRTF

In the current experiment the HRTFs are represented in one of three forms: H_{mla}, H_{ml} and H_{mlg}. In this work, all HRTFs are represented as impulse responses sampled at 48 kHz. First of all, an impulse response is made from the measurement where the first nonzero value is identified by visual inspection. All values before this 'starting point' are considered as measurement noise and replaced by zeros. The length of the initial delay constitutes the linear phase component. The impulse response from the first nonzero sample onwards is decomposed into a minimum phase and an all-pass component. A convenient method of obtaining the minimum phase component is to find the roots of the polynomial that are outside the unit circle in the z-plane. These are mirrored to their reciprocal positions and the polynomial is reconstructed. However, significant round-off errors may occur for filter orders higher than approximately 64. Consequently, homomorphic filtering employing the real cepstrum (as described in Oppenheim and Shafer [2]) is used instead to obtain the minimum phase component. The minimum phase component is convolved with the linear phase component to obtain H_{ml}.

$$H(z)_{ml} = H(z)_{minimum\ phase} \cdot H(z)_{linear\ phase} \tag{3}$$

The all-pass component can be found by dividing the original transfer function (without the initial delay) by its minimum phase component in the frequency domain. This, however, forces an all-pass component of a particular order (the same as the original transfer function). Instead, the all-pass component is constructed from the roots outside the unit circle by creating poles at reciprocal positions. Every complex conjugate pair of poles and zeros then corresponds to a single second order all-pass section and a real pole/zero pair corresponds to a first order all-pass section. In this way the all-pass is component decomposed into an integer number of first and second order all-pass sections.

H_{ml} is consecutively filtered with every first and second order section in the all-pass component to give an impulse response H_{mla}. The all-pass component as a whole can be used for filtering but its construction is complicated by the need to construct a polynomial from the roots. Round-off errors can occur if the number of cascaded all-pass sections is large, as is often the case for contralateral HRTFs. Since the methods for decomposition described here are numerically robust H_{mla} is exactly equivalent to the original HRTF.

$$H(z)_{mla} = H(z)_{minimum\ phase} \cdot H(z)_{linear\ phase} \cdot H(z)_{all-pass} \tag{4}$$

$\tau_g(f)$ for the first and second order sections in the all-pass component is accumulated to obtain the group delay of the complete all-pass component. This function is evaluated at 0 Hz to obtain the number of samples (rounded to an integer) that is added to the beginning of the impulse response of H_{ml} to obtain H_{mlg}. In this way, the all-pass component has been replaced by a pure delay.

$$H(z)_{mlg} = H(z)_{minimum\ phase} \cdot H(z)_{linear\ phase} \cdot H(z)_{\tau_g(0)} \tag{5}$$

Examples of an HRTF in its H_{mla}, H_{ml} and H_{mlg} forms are shown in Figure 3. The impulse responses shown are for the contralateral side at (90, 0), i.e. for the right ear while the sound source was positioned directly to the left of the head in the horizontal plane.

4.2 Conditions

Twelve pairs of HRTFs divided into two groups of six were selected from the database of 40 people measured at 97 directions [1]. The first group had the largest IGD_0 of the all-pass components of all the HRTFs in the database. For this group the IGD_0 was above the 30 μs threshold. The second group consisted of 'typical' HRTFs showing a small IGD_0 of the all-pass components. This group represents most HRTFs measured on the sphere around a person. In the listening experiment HRTFs in their H_{ml} and H_{mlg} form were tested against their H_{mla} form as follows:

Condition A: IGD_0 of all-pass components above threshold (H_{ml} vs. H_{mla})

The first group of 6 HRTFs pairs in this condition are in their H_{ml} form and have the largest IGD_0 determined for the all-pass component of all HRTFs in the database. The directions, (azimuth, elevation), where this occur are (-90, 0), (90, 0), (112.5, 0), (112.5, -22.5), (-112.5, -22.5) and (-90, 0). The co-ordinate system has its poles above and below the head; positive azimuths are to the left and positive elevations are in the upward direction.

Condition B: IGD_0 of all-pass components above threshold (H_{mlg} vs. H_{mla})

The HRTFs are the same as in Condition A but here they are in the form H_{mlg}. All operations are done on the data sampled at 48 kHz. Since $\tau_g(0)$ is not an integer number of samples, it was ensured that the error in interaural time in H_{mlg} was less than one sample period. The delay (IGD_0 determined from the all-pass components) added to the 6 HRTFs in Condition A turned out to be 3 samples in all cases.

Condition C: IGD_0 of all-pass components below threshold (H_{ml} vs. H_{mla})

The 6 HRTF pairs of the second group in this condition correspond to the directions (0, 0), (180, 0), (45, -45), (135, 45), (-135, -45) and (-45, 45) and the person the HRTFs were measured on was randomly selected. The HRTFs in this condition represent the majority of measured HRTFs in terms of their all-pass components. Although the HRTFs used here are in principle in the H_{ml} form, H_{mlg} is essentially the same at the sampling rate of 48 kHz, i.e. the IGD_0 from the all-pass components is 0 samples (when rounded).

Condition D: IGD_0 of all-pass components below threshold ($H_{ml\,\mathrm{perturbed}}$ vs. H_{mla})

In a 3-AFC experiment it is important not to frustrate listeners by presenting differences that are too subtle. Therefore, the differences should be clearly audible for a reasonable percentage of the presentations. No detection for Conditions B and C and only partial detection for Condition A was expected. So a set of HRTFs was created, for which high detection was expected by slightly changing the ITD of the HRTFs of the second group (same as in Condition C). The ITD was diminished by 3 samples for the (0, 0) and (180, 0) directions and by 4 samples for the (45, -45), (135, 45), (-135, -45) and (-45, 45) directions.

4.3 Stimuli

A signal was created by filtering a 650 ms ramped pink noise with an HRTF pair. A pair of HRTFs in their H_{mla} form was compared to either its H_{ml} or H_{mlg} form. Sequences of three consecutive sounds with 100 ms of silence in-between were created. The 6 possible combinations of ordering the sequence were presented twice, giving 12 presentations per HRTF pair.

The experiment was divided into two parts each containing 3 conditions: Part 1) Conditions A, C and D and Part 2) Conditions A, B and D. In each part the 6 HRTFs in the 3 conditions were randomized for every listener to give 216 presentations. These were divided into 6 listening sessions of approximately 10 minutes each. Part 1 was run with all listeners and the results were analyzed. Those listeners that detected the differences in Condition A with statistical significance continued with Part 2 of the experiment on another day.

4.4 Subjective test

Twelve listeners (6 male and 6 female) participated in the subjective test. They had normal hearing according to an audiometric test and were between the ages of 20 and 30. All listeners had participated in listening tests before but were unfamiliar with the differences tested. A listener was seated at a table in a quiet empty room. The listener wore equalized Beyerdynamic DT990 headphones and interacted with an electronic tablet showing the sketch in Figure 4.

A presentation of a sequence of three sounds was presented over the headphones. An answer was submitted by selecting one of the three areas on the bottom of the sketch with an electronic pen. The listener could repeat a presentation up to 3 times by selecting the repeat (GENTAG) button. Before the start of the experiment the listener was familiarized with the task by presenting a typical test run. No feedback on performance was given.

4.5 Statistical analysis

In a 3-AFC experiment the number of correct answers is binomially distributed. The null hypothesis is that the listener is guessing (chance performance). The probability of guessing a correct answer is 33.3%. The hypothesis of chance performance is rejected if the p-value is smaller than the significance level. For every condition in the current experiment 72 answers were recorded from each listener. Therefore, the null hypothesis is rejected at the 1 % significance level if the percentage of correct answers is greater than 46 %.

5 Results

The results of the subjective experiment are shown in Figure 5. All 12 listeners participated in the first part of the experiment that included Conditions A, C and D. Only 7 listeners met the criterion to continue in the second part of the experiment that included Conditions A, B and D. The bars show the percentages of correct answers given by each listener. The significance boundary at the 1 % level is indicated by a dashed line. Percentages of correct answers above the line mean that the null hypothesis has to be rejected, i.e. the presented differences were detectable. This is the case for some listeners in Condition A and Condition D, but not in Condition B and Condition C.

6 Discussion

An interpretation of the results of Minnaar et al. [10] lead to the criterion for the audibility of the all-pass components in HRTFs. If the interaural group delay difference at 0 Hz (IGD_0) introduced by the all-pass components of the two sides is below 30 μs, removing them will not be audible. This indicates that the auditory system will not detect an IGD_0 of less than about 1.5 samples at 48 kHz. Inspection of the HRTF measurements made on 40 people [1] shows that the IGD_0 determined from the all-pass components are generally below the threshold in which case the all-pass components can be removed without audible consequences. Therefore, the HRTFs can be represented in their H_{ml} form as seen in the results of Condition C. Since the H_{ml} forms are identical to the H_{mlg} forms for the HRTFs in Condition C, these HRTFs can also be represented in their H_{mlg} forms.

The HRTFs in Condition A are also in their H_{ml} form. However, for these HRTFs the all-pass components cannot simply be removed without introducing an audible change. The results of Condition A show that the introduced change was detected by 7 of the 12 listeners. The IGD_0 of the all-pass components in Condition A are 3 samples in all cases. This is above the 30 μs threshold explaining the audibility. The IGD_0 of the all-pass components has been added to the HRTFs in Condition A (H_{ml}) to form those in Condition B (H_{mlg}). The results of Condition B are below the 1 % significance boundary indicating that listeners were guessing.

Both in Condition B and in Condition C HRTFs in their H_{mlg} form were indistinguishable from the H_{mla} forms. Therefore, a pair of HRTFs can always be replaced by its minimum phase components and an ITD calculated as the IGD_0 of the excess phase components. This result implies that the auditory system is not sensitive to the exact structure of the phase of HRTFs and that the exact time development is not perceptible.

In a 3-AFC experiment it is important to provide a listener with a reasonable percentage of presentations where the differences are clearly audible. If the experiment is designed such that the overall result is near 33%, it cannot be determined whether a listener is unable to hear the differences or is uninterested in the experiment. This motivated the separation of the experiment into two parts and the introduction of Condition D for which detection was expected. It is clear from the results that the perturbations introduced in Condition D were small, as was the intention.

In principle, the value of the HRTF magnitude is unity and the phase is zero at 0 Hz. If this is not the case, due to, e.g., high-pass filtering of the HRTF or poor low frequency estimation, the delay used with the minimum phase components cannot be determined from IGD_0 of the excess phase components. In Figure 2 it is seen that the group delay function is smooth and approximately flat at frequencies below 1500 Hz, i.e. the phase is a linear function of frequency. This is a property of HRTFs that have been divided by the measurement at a microphone position in the middle of the head. In the HRTF-database of 40 people the group delay functions are approximately constant up to 1500 Hz and deviations from the value at 0 Hz, $\tau_g(0)$, are within 30 μs. Consequently, any group delay value in the range from 0 to 1500 Hz can be used to calculate the appropriate delay to be applied to the minimum phase components.

7 Conclusion

A listening experiment was done to test the audibility of all-pass components in HRTFs. It was shown that the all-pass components of most HRTFs can be removed without audible consequences. However, for some HRTFs their absence is detectable. For these HRTFs the all-pass components can be replaced by pure delays calculated as the group delays of the all-pass components evaluated at 0 Hz. Therefore, an HRTF pair can be represented by minimum phase transfer functions and an ITD which is the IGD_0 of the excess phase.

8 Acknowledgement

We want to thank the patient listeners that participated in the experiments. Many thanks to Jesper Sandvad for his inspiring contributions to the work and to Claus Vestergaard Skipper for his practical help during the preparation of the experiment. This work was funded by the Danish National Centre for IT Research.

9 References

[1] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen, "Head-related transfer functions of human subjects", *Journal of the Audio Engineering Society*, Vol. 43, No. 5, pp. 300-321, 1995.

[2] A. V. Oppenheim, R. W. Shafer, "Discrete time signal processing", *Prentice Hall International*, 1989.

[3] S. Mehrgardt, V. Mellert, "Transformation characteristics of the external human ear", *Journal of the Acoustical Society of America*, Vol. 61, pp.1567-1576, 1977.

[4] J-M. Jot, V. Larcher, O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony", *Proceedings of the 106th Convention of the Audio Engineering Society*, Paris, France, Preprint 3980, 1995.

[5] J. Huopaniemi, J. O. Smith, "Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters", *Proceedings of the AES 16th International Conference*, Rovaniemi, Finland, pp. 301 -312, 1999.

[6] D. J. Kistler, F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", *Journal of the Acoustical Society of America*, Vol. 91, pp. 1637-1647, 1992.

[7] J. Sandvad, D. Hammershøi, "Binaural auralization. Comparison of FIR and IIR Filter representation of HIRs", *Proceedings of the 96th Convention of the Audio Engineering Society*, Amsterdam, The Netherlands, Preprint 3862, 1994.

[8] P. J. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, H. Møller, "The interaural time difference in binaural synthesis", *Presented at the 108th Convention of the Audio Engineering Society*, Paris, France, 2000.

[9] A. Kulkarni, S. K. Isabelle, H. S. Colburn, "Sensitivity of human subjects to head-related transfer function phase spectra", *Journal of the Acoustical Society of America*, Vol. 105 (5), pp. 2821-2840, 1999.

[10] P. J. Minnaar, F. Christensen, H. Møller, S. K. Olesen, J. Plogsties, "Audibility of all-pass components in binaural synthesis", *Proceedings of the 106th Convention of the Audio Engineering Society*, Munich, Germany, Preprint 4911, 1999.

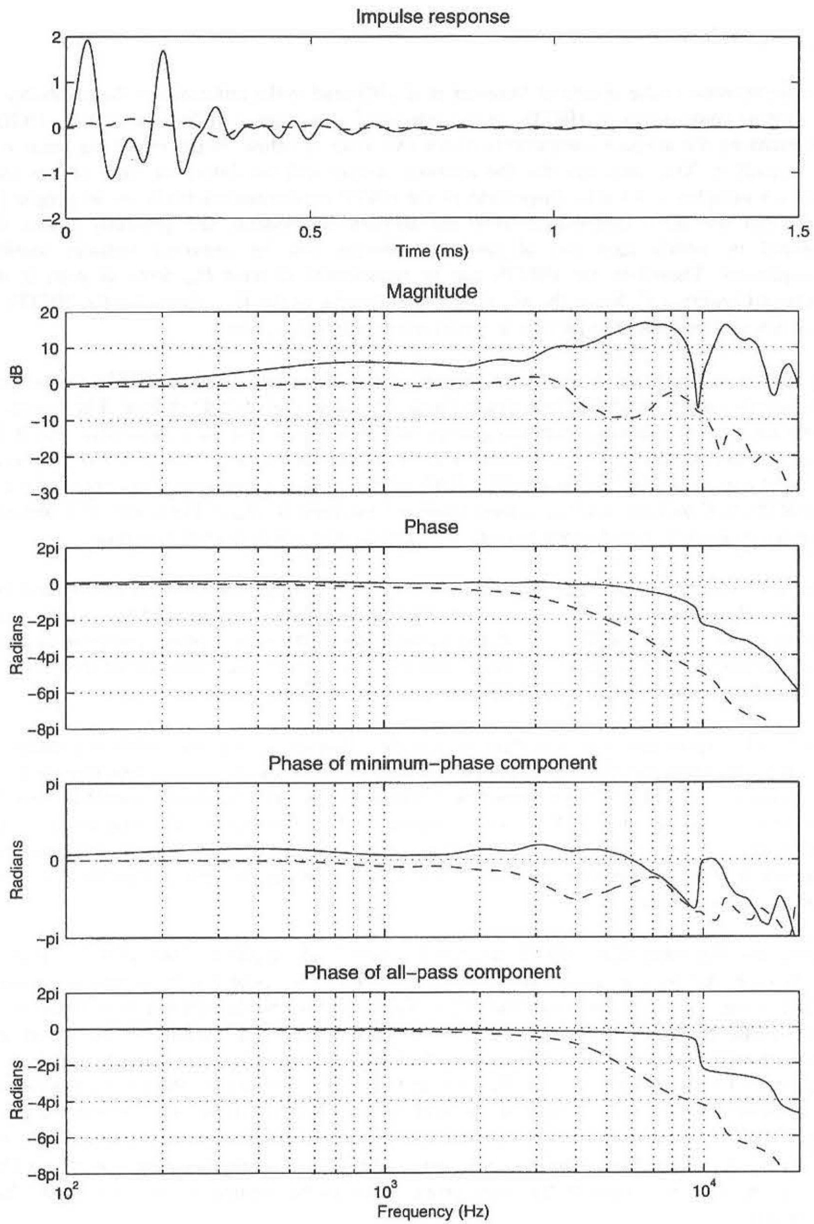

Figure 1. One set of HRTFs at (90,0) with the linear phase component (initial delay) removed, decomposed into minimum phase and all-pass components. Left ear (solid line) and right ear (dashed line).

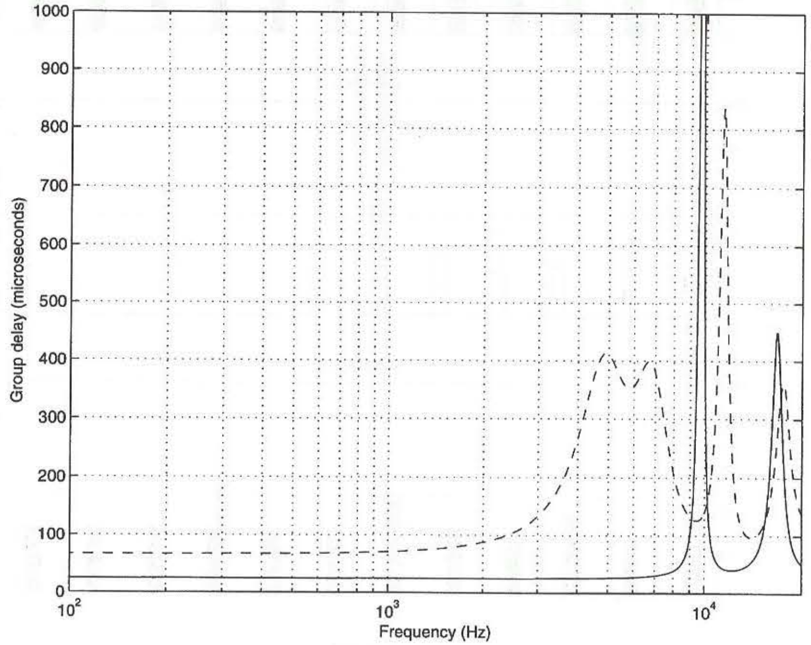

Figure 2. The group delay is calculated from the all-pass component of the set of HRTFs shown in Figure 1. Left ear (solid line) and right ear (dashed line).

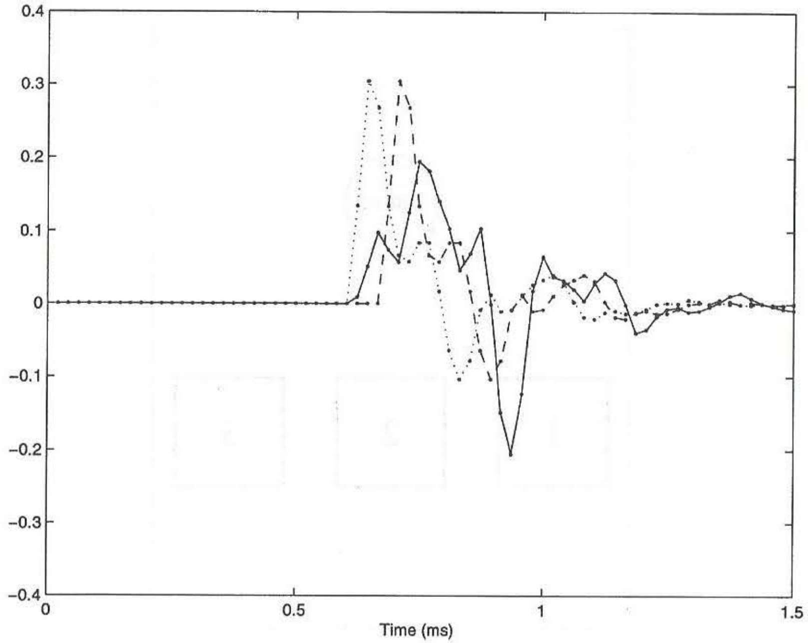

Figure 3. A contralateral HRTF at (90, 0), sampled at 48 kHz, in the three forms: a) H_{mla} (solid line), b) H_{ml} (dotted line), c) H_{mlg} (dashed line).

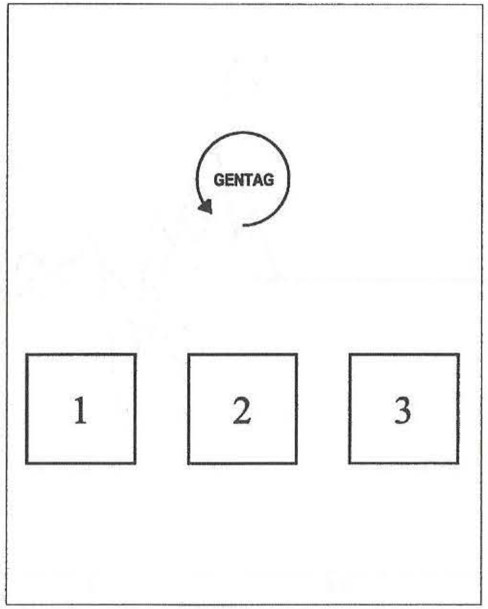

Figure 4. The interface used by listeners in the 3-AFC experiment. A presentation could be repeated up to 3 times by selecting the repeat (GENTAG) button. After that the listeners had to indicate their choice by pressing one of the numbered buttons.

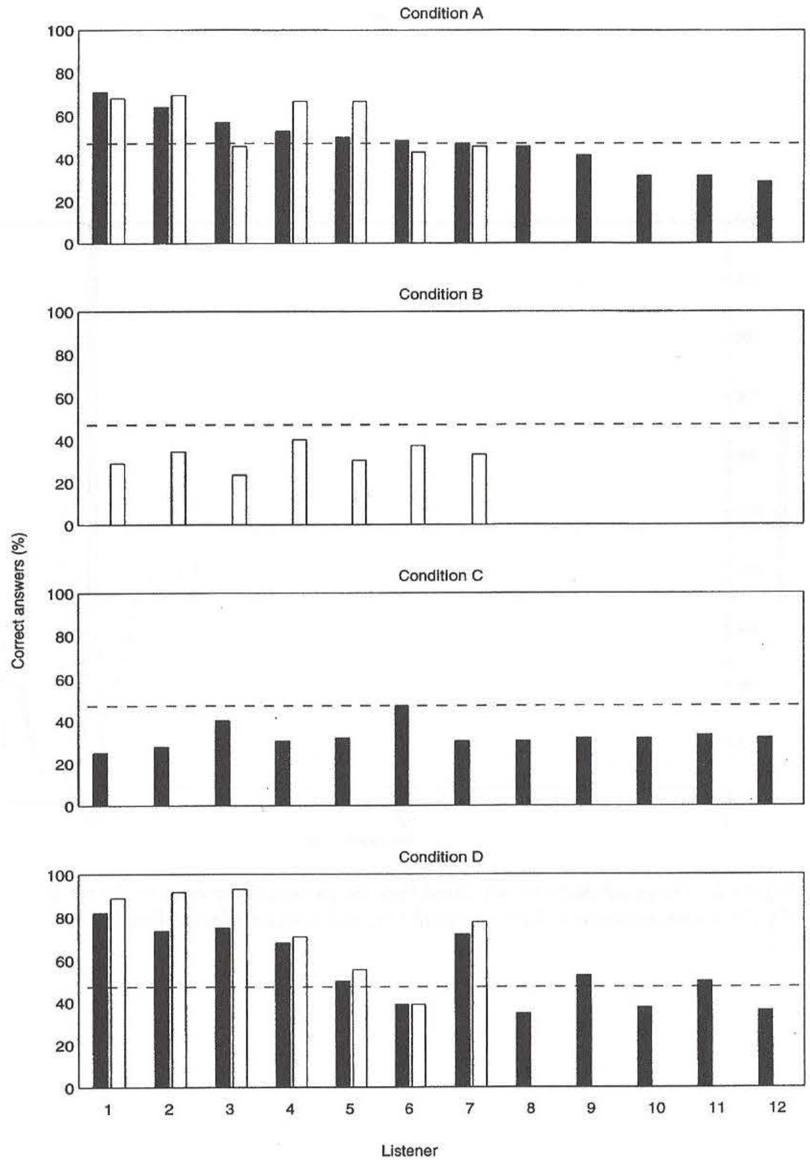

Figure 5. Results of Part 1 (black) and Part 2 (white) of the listening experiment:
Condition A: IGD_0 of all-pass component above threshold (H_{ml})
Condition B: IGD_0 of all-pass component above threshold (H_{mlg})
Condition C: IGD_0 of all-pass component below threshold ($H_{ml} = H_{mlg}$)
Condition D: IGD_0 of all-pass component below threshold ($H_{ml\ \mathrm{perturbed}}$)